

# An optimal problem for relative entropy

Fan Wang, Jun Zhu\*and Lin Zhang

*Institute of Mathematics, Hangzhou Dianzi University, Hangzhou 310018, PR China*

## Abstract

Relative entropy is an essential tool in quantum information theory. There are so many problems which are related to relative entropy. In this article, the optimal values which are defined by  $\max_{U \in U(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma)$  and  $\min_{U \in U(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma)$  for two positive definite operators  $\rho, \sigma \in \text{Pd}(\mathcal{X})$  are obtained. And the set of  $S(U\rho U^* \parallel \sigma)$  for every unitary operator  $U$  is full of the interval  $[\min_{U \in U(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma), \max_{U \in U(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma)]$ .

**Keywords:** Quantum operation; Unitary operator; Relative entropy

## 1 Introduction

In complex Euclidean spaces  $\mathcal{X}$ , one writes  $L(\mathcal{X}, \mathcal{Y})$  to refer to the collection of all linear mappings of the form  $A : \mathcal{X} \rightarrow \mathcal{Y}$ , denote  $L(\mathcal{X}, \mathcal{X})$  by  $L(\mathcal{X})$ , an operator  $A \in L(\mathcal{X})$  is normal if and only if it commutes with its adjoint:  $[A, A^*]$ , or equivalently  $AA^* = A^*A$ . An operator is *positive semidefinite*  $A \in L(\mathcal{X})$  if and only if it holds that  $A = B^*B$  for some operator  $B \in L(\mathcal{X})$ . The collection of operators is denoted  $\text{Pos}(\mathcal{X}) = \{B^*B : B \in L(\mathcal{X})\}$ . The notation  $P \geq 0$  is also used to mean that  $P$  is positive semidefinite, while  $A \geq B$  means that  $A - B$  is positive semidefinite. A positive semidefinite operator  $P \in \text{Pos}(\mathcal{X})$  is said to be positive definite if, in addition to being positive semidefinite, it is non-singular. We write  $\text{Pd}(\mathcal{X}) = \{A \in \text{Pos}(\mathcal{X}) : \det(A) \neq 0\}$  to denote the set of such operators for a given complex Euclidean space  $\mathcal{X}$ .

An unitary operator  $A \in L(\mathcal{X}, \mathcal{Y})$  is a linear isometry if it preserves the Euclidean norm-meaning that  $\|Au\| = \|u\|$  for all  $u \in \mathcal{X}$ . The condition that  $\|Au\| = \|u\|$  for all  $u \in \mathcal{X}$  is equivalent to  $AA^* = \mathbb{1}_{\mathcal{X}}$ . The collection of unitary operator is denoted  $U(\mathcal{X}, \mathcal{Y}) = \{A \in L(\mathcal{X}, \mathcal{Y}) : AA^* = \mathbb{1}_{\mathcal{X}}\}$ .

---

\*E-mail: junzhu@yahoo.cn

**Theorem 1.1** (Spectral Theorem([1])). *Let  $\mathcal{X}$  be a complex Euclidean space, let  $A \in \mathcal{L}(\mathcal{X})$  be a normal operator, and assume that the distinct eigenvalues of  $A$  are  $\lambda_1, \dots, \lambda_k$ . Then there exist an orthogonal basis  $\{x_1, x_2, \dots, x_n\}$  of  $\mathcal{X}$ , such that*

$$A = \sum_{i=1}^k \lambda_i x_i x_i^*.$$

von Neumann entropy and relative entropy are powerful tools in quantum information theory. The quantum relative entropy(or relative von Neumann entropy) is indispensable as a tool for the von Neumann entropy.

For two positive definite operators  $P, Q \in \text{Pd}(\mathcal{X})$ , then the von Neumann entropy of  $P$  is defined by

$$S(P) \stackrel{\text{def}}{=} \text{Tr}(P \log(P)),$$

the quantum relative entropy between  $P$  and  $Q$  is defined by:

$$S(P \parallel Q) = \text{Tr}(P \log(P)) - \text{Tr}(P \log(Q)),$$

we take the base 2 logarithm of corresponding eigenvalues. Specifically, when  $\text{im}(P) \subsetneq \text{ker} Q$ , we define  $S(P \parallel Q) = \infty$ .

Let  $B = [b_{ij}]$  be a  $N \times N$  bi-stochastic matrix, that is,  $b_{ij} \geq 0$ , and

$$\sum_{i=1}^k b_{ij} = \sum_{j=1}^k b_{ij} = 1,$$

for each  $i, j = 1, \dots, N$ . We only consider finite-dimensional complex Hilbert spaces, let  $\mathbf{r}, \mathbf{s} \in \mathbb{R}^n$ ,  $\mathbf{r}$  is *majorised* by  $\mathbf{s}$ , written as  $\mathbf{r} \prec \mathbf{s}$ , if

$$\sum_{i=1}^k r_i \leq \sum_{j=1}^k s_i, \quad 1 \leq k \leq n-1$$

and

$$\sum_{i=1}^n r_i = \sum_{j=1}^n s_i.$$

## 2 Description of the problems and main results

For two positive definite operators  $P, Q \in \text{Pd}(\mathcal{X})$ , here  $\dim(\mathcal{X}_d) = d$ . We try to find the optimal values which are defined by

$$\max_{U \in \mathcal{U}(\mathcal{X}_d)} S(U \rho U^* \parallel \sigma) \quad \text{and} \quad \min_{U \in \mathcal{U}(\mathcal{X}_d)} S(U \rho U^* \parallel \sigma)$$

respectively. Since

$$\max_{U \in U(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma) = -S(\rho) - \min_{U \in U(\mathcal{X}_d)} \text{Tr}(U\rho U^* \log \sigma)$$

and

$$\min_{U \in U(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma) = -S(\rho) - \max_{U \in U(\mathcal{X}_d)} \text{Tr}(U\rho U^* \log \sigma)$$

Thus the above optimal problems are reduced to the followings:

$$\max_{U \in U(\mathcal{X}_d)} \text{Tr}(U\rho U^* \log \sigma) \quad \text{and} \quad \min_{U \in U(\mathcal{X}_d)} \text{Tr}(U\rho U^* \log \sigma)$$

Firstly, we assume that  $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$  and  $\sigma$  is of full rank. By the spectral decomposition, we have that

$$\sigma = \sum_j \lambda_j^\downarrow(\sigma) |y_j\rangle \langle y_j| = \sum_j \lambda_j^\uparrow(\sigma) |z_j\rangle \langle z_j|.$$

For an orthonormal basis  $\{|y_j\rangle\}$  of  $\mathcal{X}_d$  and  $|y_k\rangle = |z_{d+1-k}\rangle$  for  $k \in [d] = \{1, \dots, d\}$ . Here  $\lambda_j^\downarrow(\sigma)$  stands for eigenvalues are arranged in decreasing order and  $\lambda_j^\uparrow(\sigma)$  stands for eigenvalues are arranged in increasing order. If  $[\rho, \sigma] \neq 0$ , from the spectral decomposition of  $\rho$ :

$$\rho = \sum_j \lambda_j^\downarrow(\rho) |x_j\rangle \langle x_j|,$$

where  $\{|x_j\rangle\}$  is an orthonormal basis of  $\mathcal{X}_d$ , we know that there exist two unitary operators  $U, V : U = \sum_j |y_j\rangle \langle x_j|$  and  $V = \sum_j |z_j\rangle \langle x_j|$ , such that  $U|x_j\rangle = |y_j\rangle$  for each  $j$ , respectively.

Now

$$\text{Tr}(U\rho U^* \log \sigma) = \sum_j \lambda_j^\downarrow(\rho) \log \lambda_j^\downarrow(\sigma)$$

and

$$\text{Tr}(V\rho V^* \log \sigma) = \sum_j \lambda_j^\downarrow(\rho) \log \lambda_j^\uparrow(\sigma).$$

In mathematics, the rearrangement inequality states that

$$x_n y_1 + \dots + x_1 y_n \leq x_{\pi(1)} y_1 + \dots + x_{\pi(n)} y_n \leq x_1 y_1 + \dots + x_n y_n \quad (2.1)$$

for every choice of real numbers

$$x_1 \leq \dots \leq x_n \quad \text{and} \quad y_1 \leq \dots \leq y_n,$$

and every permutation  $x_{\pi(1)}, \dots, x_{\pi(n)}$  of  $x_1 \leq \dots \leq x_n$ . If the number are different, meaning that

$$x_1 < \dots < x_n \quad \text{and} \quad y_1 < \dots < y_n,$$

then the lower bound is attained only for the permutation which reverses the order, i.e.  $\pi(i) = n - i + 1$  for all  $i = 1, \dots, n$ .

Note that the rearrangement inequality makes no assumptions on the signs of the real number.

**Lemma 2.1.** Let  $\mathbf{x}^\downarrow, \mathbf{y}^\downarrow, \mathbf{u}^\downarrow \in \mathbb{R}^d$ , all having their coordinates in decreasing order. If  $\mathbf{y}^\downarrow \prec \mathbf{x}^\downarrow$ , i.e.  $\mathbf{y}^\downarrow$  is majorized by  $\mathbf{x}^\downarrow$ , then

$$\langle \mathbf{u}^\downarrow, \mathbf{y}^\downarrow \rangle \leq \langle \mathbf{u}^\downarrow, \mathbf{x}^\downarrow \rangle.$$

*Proof.* For the completeness, we give the detailed proof. We make an induction argument in the following.

We denote the  $\text{Tr}(\mathbf{x}^\downarrow) = \text{Tr}(\mathbf{y}^\downarrow) = s$ , i.e. ,  $s = \sum_{j=1}^d x_j^\downarrow = \sum_{j=1}^d y_j^\downarrow$ . If  $d = 2$ , by a simple algebra, we have

$$\langle \mathbf{u}^\downarrow, \mathbf{y}^\downarrow \rangle \leq \langle \mathbf{u}^\downarrow, \mathbf{x}^\downarrow \rangle.$$

In fact,

$$\sum_{j=1}^2 x_j^\downarrow u_j^\downarrow - \sum_{j=1}^2 x_j^\downarrow u_j^\downarrow = (su_1^\downarrow + x_2^\downarrow(u_2^\downarrow - (su_1^\downarrow))) - (su_1^\downarrow + y_2^\downarrow(u_2^\downarrow - u_1^\downarrow)) = (x_2^\downarrow - y_2^\downarrow)(u_2^\downarrow - u_1^\downarrow) \geq 0$$

Where the last line in the above inequality holds because  $\mathbf{y}^\downarrow$  is majorized by  $\mathbf{x}^\downarrow$ , i.e. ,

$$x_1^\downarrow \geq y_1^\downarrow, \quad s = \sum_j x_j^\downarrow = \sum_j y_j^\downarrow.$$

Next we suppose that the inequality holds for  $d - 1$ . Now we show that the inequality holds for  $d$ .

$$\sum_{j=1}^d x_j^\downarrow u_j^\downarrow - \sum_{j=1}^d x_j^\downarrow u_j^\downarrow = \sum_{j=1}^{d-1} (x_j^\downarrow - y_j^\downarrow)(u_j^\downarrow - u_d^\downarrow) \geq (u_{d-1}^\downarrow - u_d^\downarrow) \sum_{j=1}^{d-1} (x_j^\downarrow - y_j^\downarrow) \geq 0$$

The proof is completed. □

**Remark 2.2.** The result obtained in Lemma 2.1 can be reformulated as: For two fixed  $d$ -dimensional real vector  $\mathbf{x}^\downarrow, \mathbf{u}^\downarrow \in \mathbb{R}^d$ , we have

$$\max_B \langle \mathbf{u}^\downarrow | B | \mathbf{x}^\downarrow \rangle = \langle \mathbf{u}^\downarrow, \mathbf{x}^\downarrow \rangle,$$

where the above optimization is taken over the set of all bi-stochastic matrices of size  $d \times d$ .

By employing the rearrangement inequality, we have that, for any permutation  $\pi \in S_d$ ,

$$\sum_j \lambda_j^\downarrow(\rho) \log \lambda_j^\downarrow(\sigma) \geq \sum_j \lambda_j^\downarrow(\rho) \log \lambda_{\pi(j)}^\downarrow(\sigma) \geq \sum_j \lambda_j^\downarrow(\rho) \log \lambda_j^\uparrow(\sigma).$$

Thus from the above discussion, it is seen that the above questions are reduced to the case where  $[\rho, \sigma] = 0$  and  $\sigma > 0$ .

Now we can make an assumption:

$$\rho = \sum_j \lambda_j^\downarrow(\rho) |j\rangle \langle j| \quad \text{and} \quad \sigma = \sum_j \lambda_j^\downarrow(\sigma) |j\rangle \langle j|,$$

where  $\lambda_j^\downarrow(\sigma) > 0$  for each  $j$ . We still show that

$$\max_{U \in U(\mathcal{X}_d)} \text{Tr}(U \rho U^* \log \sigma) = \sum_j \lambda_j^\downarrow(\rho) \log \lambda_j^\downarrow(\sigma) \quad (2.2)$$

$$\min_{U \in U(\mathcal{X}_d)} \text{Tr}(U \rho U^* \log \sigma) = \sum_j \lambda_j^\downarrow(\rho) \log \lambda_j^\uparrow(\sigma) \quad (2.3)$$

Now for any unitary operator  $W \in U(\mathcal{X}_d)$ ,

$$\text{Tr}(W \rho W^* \log \sigma) = \sum_{i,j} \lambda_i^\downarrow(\rho) \log \lambda_j^\downarrow(\sigma) | \langle j | W | i \rangle |^2.$$

Setting  $D = [| \langle j | W | i \rangle |^2]$ , we see that  $D$  is a bi-stochastic matrix since  $D = W \circ \overline{W}$ , where  $\circ$  is Schur product.

Now we have  $\text{Tr}(W \rho W^* \log \sigma) = \sum_j \lambda_j^\downarrow(\rho) \alpha_j$ , where  $\alpha_j = \sum_i | \langle j | W | i \rangle |^2 \lambda_i^\downarrow(\rho) = \sum_i D_{ij} \lambda_i^\downarrow(\rho)$ .

This indicates that

$$\alpha = D \lambda^\downarrow(\rho),$$

where

$$\alpha \stackrel{\text{def}}{=} [\alpha_1, \dots, \alpha_d]^\top \quad \text{and} \quad \lambda^\downarrow(\rho) \stackrel{\text{def}}{=} [\lambda_1^\downarrow(\rho), \dots, \lambda_d^\downarrow(\rho)]^\top,$$

which implies that

$$\alpha \prec \lambda^\downarrow(\rho).$$

Therefore

$$\text{Tr}(W \rho W^* \log \sigma) = \langle \log \lambda^\downarrow(\rho) | D | \lambda^\downarrow(\rho) \rangle,$$

where  $\log \lambda^\downarrow(\rho) \stackrel{\text{def}}{=} [\log \lambda_1^\downarrow(\rho), \dots, \log \lambda_d^\downarrow(\rho)]^\top$ .

In what follows, we show that

$$\max_{U \in U(\mathcal{X}_d)} \text{Tr}(W\rho W^* \log \sigma) = \max_D \langle \log \lambda^\downarrow(\rho) | D | \lambda^\downarrow(\rho) \rangle.$$

It follows from Lemma 2.1 that

$$\max_D \langle \log \lambda^\downarrow(\rho) | D | \lambda^\downarrow(\rho) \rangle = \langle \log \lambda^\downarrow(\rho), \lambda^\downarrow(\rho) \rangle = \sum_{j=1}^d \lambda_j^\downarrow(\rho) \log \lambda_j^\downarrow(\rho).$$

Thus we are arriving at the following conclusion:

**Theorem 2.3.** *Let  $\mathcal{X}$  be a complex Euclidean spaces and assume that  $\rho, \sigma \in \text{Pd}(\mathcal{X})$  are positive definite operators on  $\mathcal{X}$ . Then*

$$\max_{U \in U(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma) = -S(\rho) - \langle \log \lambda^\downarrow(\rho), \lambda^\downarrow(\rho) \rangle.$$

Similarly, we have

**Theorem 2.4.** *Let  $\mathcal{X}$  be a complex Euclidean spaces and assume that  $\rho, \sigma \in \text{Pd}(\mathcal{X})$  are positive definite operators on  $\mathcal{X}$ . Then*

$$\min_{U \in U(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma) = -S(\rho) - \langle \log \lambda^\downarrow(\rho), \lambda^\uparrow(\rho) \rangle.$$

### 3 Further result

In this section, the value range of  $S(U\rho U^* \parallel \sigma)$  is discussed. We define the following sets

$$M(\rho, \sigma) \stackrel{\text{def}}{=} \{S(U\rho U^* \parallel \sigma) : U \in U(\mathcal{X}_d)\}, \quad \tilde{M}(\rho, \sigma) \stackrel{\text{def}}{=} \{\text{Tr}(U\rho U^* \sigma) : U \in U(\mathcal{X}_d)\}$$

If the set  $M(\rho, \sigma)$  is the interval

$$[\min_{U \in U(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma), \max_{U \in U(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma)]$$

it is equivalent to the following:

$$\tilde{M} = [\min_{U \in U(\mathcal{X}_d)} \text{Tr}(U\rho U^* \sigma), \max_{U \in U(\mathcal{X}_d)} \text{Tr}(U\rho U^* \sigma)]$$

**Theorem 3.1.** *Let  $\mathcal{X}$  be a complex Euclidean spaces and assume that  $\rho, \sigma \in \text{Pd}(\mathcal{X})$  are positive definite operators on  $\mathcal{X}$ . If we define the following sets*

$$\tilde{M}(\rho, \sigma) \stackrel{\text{def}}{=} \{\text{Tr}(U\rho U^* \sigma) : U \in U(\mathcal{X}_d)\},$$

then

$$\tilde{M}(\rho, \sigma) = [\min_{U \in U(\mathcal{X}_d)} \text{Tr}(U\rho U^* \sigma), \max_{U \in U(\mathcal{X}_d)} \text{Tr}(U\rho U^* \sigma)]$$

*Proof.* This problem is equivalent to

$$\tilde{M}(\rho, \sigma) = [\min_{U \in U(\mathcal{X}_d)} \text{Tr}(U\rho U^* \sigma), \max_{U \in U(\mathcal{X}_d)} \text{Tr}(U\rho U^* \sigma)] = [\sum_{i=1}^n \lambda_i^\downarrow(\rho) \log \lambda_i^\uparrow(\sigma), \sum_{i=1}^n \lambda_i^\downarrow(\rho) \log \lambda_i^\downarrow(\sigma)].$$

If  $d = n$  we assume that  $\lambda^\downarrow(\rho) = (\lambda_1, \dots, \lambda_n)$  and  $\lambda^\downarrow(\sigma) = (\mu_1, \dots, \mu_n)$ . Then the problem is equivalent to

$$\tilde{M}(\rho, \sigma) = [\sum_{i=1}^n \lambda_i \mu_{n-i+1}, \sum_{i=1}^n \lambda_i \mu_i].$$

We make an induction argument in the following. And we only need consider the diagonal form of  $\rho$  and  $\sigma$  for any unitary, what's more, the value on diagonal are arranged in decreasing order.

Firstly, if  $d = 2$ , the problem is equivalent to  $\tilde{M}(\rho, \sigma) = [\lambda_1 \mu_2 + \lambda_2 \mu_1, \lambda_1 \mu_1 + \lambda_2 \mu_2]$  for  $\lambda_1, \lambda_2$  and  $\mu_1, \mu_2$  are eigenvalues of  $\rho, \sigma \in \text{Pd}(\mathcal{X})$ , respectively. For any  $k \in [0, 1]$ , there exists unitary  $U$ , such that

$$U\rho U^* = \begin{bmatrix} k\lambda_1 + (1-k)\lambda_2 & \sqrt{k}\sqrt{1-k}(\lambda_1 - \lambda_2) \\ \sqrt{k}\sqrt{1-k}(\lambda_1 - \lambda_2) & k\lambda_2 + (1-k)\lambda_1 \end{bmatrix},$$

where

$$U = \begin{bmatrix} \sqrt{k} & \sqrt{1-k} \\ \sqrt{1-k} & -\sqrt{k} \end{bmatrix}, k \in [0, 1].$$

Then

$$\begin{aligned} \text{Tr}(U\rho U^* \sigma) &= [(1-k)\lambda_2 + k\lambda_1]\mu_1 + [k\lambda_2 + (1-k)\lambda_1]\mu_2 \\ &= (1-k)(\lambda_1 \mu_2 + \lambda_2 \mu_1) + k(\lambda_1 \mu_1 + \lambda_2 \mu_2). \end{aligned}$$

It is equivalent to  $\text{Tr}(U\rho U^* \sigma)$  is convex combination of  $\lambda_1 \mu_2 + \lambda_2 \mu_1$  and  $\lambda_1 \mu_1 + \lambda_2 \mu_2$ . Therefore, when  $d = 2$ ,  $\tilde{M}(\rho, \sigma) = [\lambda_1 \mu_2 + \lambda_2 \mu_1, \lambda_1 \mu_1 + \lambda_2 \mu_2]$ .

Next we suppose that the conclusion holds for  $d = n - 1$ , i.e. if  $d = n - 1$ ,

$$\tilde{M}(\rho, \sigma) = [\sum_{i=1}^{n-1} \lambda_i^\downarrow(\rho) \log \lambda_i^\uparrow(\sigma), \sum_{i=1}^{n-1} \lambda_i^\downarrow(\rho) \log \lambda_i^\downarrow(\sigma)] = [\sum_{i=1}^{n-1} \lambda_i \mu_{n-i}, \sum_{i=1}^{n-1} \lambda_i \mu_i] \quad (3.1)$$

where  $\lambda^\downarrow(\rho) = (\lambda_1, \dots, \lambda_{n-1})$  and  $\lambda^\downarrow(\sigma) = (\mu_1, \dots, \mu_{n-1})$ .

Now we should to prove the conclusion holds for  $d = n$ .

From (2.1), we know that

$$\begin{aligned}
\sum_{i=1}^n \lambda_i^\downarrow(\rho) \log \lambda_i^\downarrow(\sigma) &= \lambda_1 \mu_1 + \lambda_2 \mu_2 + \lambda_3 \mu_3 + \dots + \lambda_n \mu_n \\
&\geq \lambda_2 \mu_1 + \lambda_1 \mu_2 + \lambda_3 \mu_3 + \dots + \lambda_n \mu_n \\
&\geq \lambda_2 \mu_1 + \lambda_3 \mu_2 + \lambda_1 \mu_3 + \dots + \lambda_n \mu_n \\
&\geq \dots \\
&\geq \lambda_2 \mu_1 + \lambda_3 \mu_2 + \lambda_4 \mu_3 + \dots + \lambda_1 \mu_{n-1} + \lambda_n \mu_n \\
&\geq \lambda_2 \mu_1 + \lambda_3 \mu_2 + \lambda_4 \mu_3 + \dots + \lambda_n \mu_{n-1} + \lambda_1 \mu_n
\end{aligned}$$

For the first inequality,

$$\begin{aligned}
&\lambda_1 \mu_1 + \lambda_2 \mu_2 + \lambda_3 \mu_3 + \dots + \lambda_n \mu_n \\
&\geq \lambda_2 \mu_1 + \lambda_1 \mu_2 + \lambda_3 \mu_3 + \dots + \lambda_n \mu_n
\end{aligned}$$

there exists an unitary  $U$ , for any  $k \in [0, 1]$ , such that

$$U \rho U^* = \begin{bmatrix} k\lambda_1 + (1-k)\lambda_2 & \sqrt{k}\sqrt{1-k}(\lambda_1 - \lambda_2) & 0 & 0 & \dots & 0 \\ \sqrt{k}\sqrt{1-k}(\lambda_1 - \lambda_2) & k\lambda_2 + (1-k)\lambda_1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & 0 & \dots & 0 \\ 0 & 0 & 0 & \lambda_4 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix},$$

where

$$U = \begin{bmatrix} \sqrt{k} & \sqrt{1-k} & 0 \\ \sqrt{1-k} & -\sqrt{k} & 0 \\ 0 & 0 & E \end{bmatrix},$$

and  $E$  is Unit matrix.

This is equivalent to  $d = 2$ , thus  $\text{Tr}(U \rho U^* \sigma)$  be full of the interval  $[\lambda_1 \mu_2 + \lambda_2 \mu_1 + \sum_{i=3}^n \lambda_i \mu_i, \lambda_1 \mu_1 + \lambda_2 \mu_2 + \sum_{i=3}^n \lambda_i \mu_i]$ . Therefore  $\text{Tr}(U \rho U^* \sigma)$  be full of the interval

$$\left[ \sum_{i=1}^{k-1} \lambda_{i+1} \mu_i + \lambda_1 \mu_k + \lambda_{k+1} \mu_{k+1} + \sum_{i=k+2}^n \lambda_i \mu_i, \sum_{i=1}^{k-1} \lambda_{i+1} \mu_i + \lambda_{k+1} \mu_k + \lambda_1 \mu_{k+1} + \sum_{i=k+2}^n \lambda_i \mu_i \right]$$

where  $k \in \{1, \dots, n\}$ .



Thus  $\text{Tr}(U\rho U^*\sigma)$  be full of the interval

$$\left[ \sum_{i=1}^{n-1} \lambda_{i+1} \mu_{n-i} + \lambda_1 \mu_n, \sum_{i=1}^{n-1} \lambda_{i+1} \mu_i + \lambda_1 \mu_n \right] \quad (3.2)$$

From (3.1) we know that if  $d = n - 1$ , for  $\lambda^\downarrow(\rho) = (\lambda_2, \dots, \lambda_n)$  and  $\lambda^\downarrow(\sigma) = (\mu_1, \dots, \mu_{n-1})$ , we have  $\tilde{M}(\rho, \sigma) = \left[ \sum_{i=1}^{n-1} \lambda_{i+1} \mu_{n-i}, \sum_{i=1}^{n-1} \lambda_{i+1} \mu_i \right]$ . Then when  $d = n$ ,  $\text{Tr}(U\rho U^*\sigma)$  be full of the interval  $\left[ \sum_{i=1}^{n-1} \lambda_{i+1} \mu_{n-i} + \lambda_1 \mu_n, \sum_{i=1}^{n-1} \lambda_{i+1} \mu_i + \lambda_1 \mu_n \right]$ .

According to (3.1) and (3.2),  $\text{Tr}(U\rho U^*\sigma)$  be full of the interval

$$\left[ \sum_{i=1}^{n-1} \lambda_{i+1} \mu_{n-i} + \lambda_1 \mu_n, \sum_{i=1}^{n-1} \lambda_{i+1} \mu_i + \lambda_1 \mu_n \right] \cup \left[ \sum_{i=1}^{n-1} \lambda_{i+1} \mu_{n-i} + \lambda_1 \mu_n, \sum_{i=1}^{n-1} \lambda_{i+1} \mu_i + \lambda_1 \mu_n \right]$$

Therefore  $\text{Tr}(U\rho U^*\sigma)$  be full of the interval

$$\left[ \sum_{i=1}^{n-1} \lambda_{i+1} \mu_{n-i} + \lambda_1 \mu_n, \sum_{i=1}^{n-1} \lambda_{i+1} \mu_i + \lambda_1 \mu_n \right] = \left[ \sum_{i=1}^n \lambda_i \mu_{n-i+1}, \sum_{i=1}^n \lambda_i \mu_i \right]$$

According to (2.2) and (2.3),

$$\max_{U \in \mathcal{U}(\mathcal{X}_d)} \text{Tr}(U\rho U^*\sigma) = \sum_{j=1}^n \lambda_j^\downarrow(\rho) \lambda_j^\downarrow(\sigma) = \sum_{i=1}^n \lambda_i \mu_i$$

and

$$\min_{U \in \mathcal{U}(\mathcal{X}_d)} \text{Tr}(U\rho U^*\sigma) = \sum_{j=1}^n \lambda_j^\downarrow(\rho) \lambda_j^\uparrow(\sigma) = \sum_{i=1}^n \lambda_i \mu_{n-i+1}.$$

Therefore  $\tilde{M}(\rho, \sigma) = \left[ \sum_{i=1}^n \lambda_i \mu_{n-i+1}, \sum_{i=1}^n \lambda_i \mu_i \right]$ . The proof is completed.  $\square$

Therefore we have

$$\begin{aligned} M(\rho, \sigma) &= \left[ \min_{U \in \mathcal{U}(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma), \max_{U \in \mathcal{U}(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma) \right] \\ &= \left[ -S(\rho) - \langle \lambda^\downarrow(\rho), \log \lambda^\downarrow(\rho) \rangle, -S(\rho) - \langle \lambda^\downarrow(\rho), \log \lambda^\uparrow(\rho) \rangle \right]. \end{aligned}$$

Then  $S(U\rho U^* \parallel \sigma)$  be full of the interval  $\left[ \min_{U \in \mathcal{U}(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma), \max_{U \in \mathcal{U}(\mathcal{X}_d)} S(U\rho U^* \parallel \sigma) \right]$ .

## 4 A generalization

In fact, the above two optimal problems can be easily generalized as follows:

Given two operators  $\rho, \sigma \in \text{Pd}(\mathcal{X})$  are positive definite operators on  $\mathcal{X}$ . If  $\omega \prec \rho$ , then

$$\max/\min\{S(\omega \parallel \sigma) : \omega \prec \rho\} = \max/\min\{S(\Phi(\rho) \parallel \sigma) : \Phi \text{ bi-stochastic}\}$$

Thus we have

$$\max_{\Phi \text{ bi-stochastic}} S(\Phi(\rho) \parallel \sigma) = -S(\rho) - \langle \lambda^\downarrow(\rho), \log \lambda^\uparrow(\rho) \rangle$$

and

$$\min_{\Phi \text{ bi-stochastic}} S(\Phi(\rho) \parallel \sigma) = -S(\rho) - \langle \lambda^\downarrow(\rho), \log \lambda^\downarrow(\rho) \rangle.$$

If we denote

$$M(\rho, \sigma) \stackrel{\text{def}}{=} \{S(\Phi(\rho) \parallel \sigma) : \Phi \text{ bi-stochastic}\}$$

Similarly, we have

$$M(\rho, \sigma) = [-S(\rho) - \langle \lambda^\downarrow(\rho), \log \lambda^\downarrow(\rho) \rangle, -S(\rho) - \langle \lambda^\downarrow(\rho), \log \lambda^\uparrow(\rho) \rangle].$$

## Acknowledgement

The author would like to thank the anonymous referee(s) for valuable comments and to thank \*\*\* for pointing out some misprints appearing in this paper. The author is also grateful for funding from Hangzhou Dianzi University .

## References

- [1] John Watrous, Theory of Quantum Information[R]. Canada ,University of Waterloo, 56-77 (2008).
- [2] T. Ando, Majorization doubly stochastic matrices and comparison of eigenvalues, Linear Algebra and its Applications, **118**: 163-248 (1989).